



City Research Online

City, University of London Institutional Repository

Citation: Chen, S., Andrienko, N., Andrienko, G., Adilova, L., Barlet, J., Kindermann, J., Nguyen, P. H., Thonnard, O. & Turkay, C. (2019). LDA Ensembles for Interactive Exploration and Categorization of Behaviors. IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2019.2904069

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/21875/>

Link to published version: <https://doi.org/10.1109/TVCG.2019.2904069>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

LDA Ensembles for Interactive Exploration and Categorization of Behaviors

Siming Chen, Natalia Andrienko, Gennady Andrienko, Linara Adilova, Jeremie Barlet, Jörg Kindermann, Phong H. Nguyen, Olivier Thonnard, and Cagatay Turkey

Abstract—We define **behavior** as a set of **actions** performed by some actor during a period of time. We consider the problem of analyzing a large collection of behaviors by multiple actors, more specifically, identifying typical behaviors and spotting anomalous behaviors. We propose an approach leveraging topic modeling techniques – LDA (Latent Dirichlet Allocation) Ensembles – to represent categories of typical behaviors by topics that are obtained through topic modeling a behavior collection. When such methods are applied to text in natural languages, the quality of the extracted topics are usually judged based on the semantic relatedness of the terms pertinent to the topics. This criterion, however, is not necessarily applicable to topics extracted from non-textual data, such as action sets, since relationships between actions may not be obvious. We have developed a suite of visual and interactive techniques supporting the construction of an appropriate combination of topics based on other criteria, such as distinctiveness and coverage of the behavior set. Two case studies on analyzing operation behaviors in the security management system and visiting behaviors in an amusement park, and the expert evaluation of the first case study demonstrate the effectiveness of our approach.

Index Terms—LDA, Visual Analytics, User Behavior.

1 INTRODUCTION

WE use the term *behavior*, or *behavior instance*, to refer to a set of *actions* performed by some actor during a period of time. With increasing data availability, behavior analysis attracts growing attention in areas such as e-commerce, computer security, health care, social media etc [1], [2]. Behavior patterns are usually complex, ill-defined and diverse. Categorization, which involves identifying and characterizing existing types (categories) of behaviors, is an important task for understanding and predicting behaviors of customers, users, or patients, as well as for spotting atypical (and possibly anomalous) behaviors. To give a simple example, consider the analysis of activity logs from an e-commerce webpage. Here, the *actions* may be “enter a search text”, “push the search button”, “scroll down”, etc., and a *behavior* where these actions appear together can be categorized as “product search”.

Our paper first investigates the suitability of topic modeling for identifying behavior types in a large set of activity logs, i.e., recorded sequences of actions performed by some *actors*, such as users of a software tool, visitors of a venue, and players of a game. This investigation was accomplished through a study conducted by an interdisciplinary team consisting of visual analytics researchers and computer security experts working on a dataset of logged sessions recording the usage of a computer application. We then further tested our approach through a synthetic dataset provided for VAST Challenge 2015. In both studies, we aim to answer the following question:

By considering a set of possible actions as a “vocabulary” and action logs as “texts” built from this vocabulary, can “topics” extracted by topic modeling algorithms be considered as representations of behavior types?

Our study allowed us to answer this question positively, but it showed a need in tools for interactive exploration and comparison of extracted “topics” and selection of a sufficient subset of representative “topics”. Thus, in this paper, we develop such tools as shown in Fig. 1. Besides, we address the problem of representing the behavior topics visually in a way that enables judgment of their quality and comparison between them. The problem is explained as follows.

When topic modeling is applied to texts in natural languages, extracted topics may be represented to the analyst as combinations of topic-relevant words. For example, words that have significantly higher probabilities of occurrence in texts associated with these topics than in texts associated with other topics. Since an analyst can interpret the meaning of the words and judge whether the word combinations make sense, the assessment of the topic quality and usefulness is straightforward. However, when the “words” are action labels, the meaning of combinations of such labels may not be clear to the analyst. Hence, it may be impossible to judge the topic quality (i.e., whether it represents a meaningful behavior type) just by looking at the list of actions that have the highest probabilities. For example, when a topic extracted from a set of text documents is represented by keywords “big data, visualization, machine learning, cloud computing”, it can be easily judged as a good topic that concerns computer science and data analysis techniques. However, if a behavior topic contains “search user, open table, close dialog”, it may be hard to judge its quality.

Therefore, there is a need for an approach to enable examination of the quality of a set of behavior topics

- S. Chen, N. Andrienko, G. Andrienko, L. Adilova, and J. Kindermann are with Fraunhofer Institute IAIS, Germany. S. Chen is also with University of Bonn, Germany. E-mail: siming.chen@iais.fraunhofer.de
- N. Andrienko, G. Andrienko, P.H. Phong, C. Turkey are with City, University of London, UK.
- O. Thonnard and J. Barlet are with Amadeus, France.

Manuscript received April 19, 2005; revised August 26, 2015.

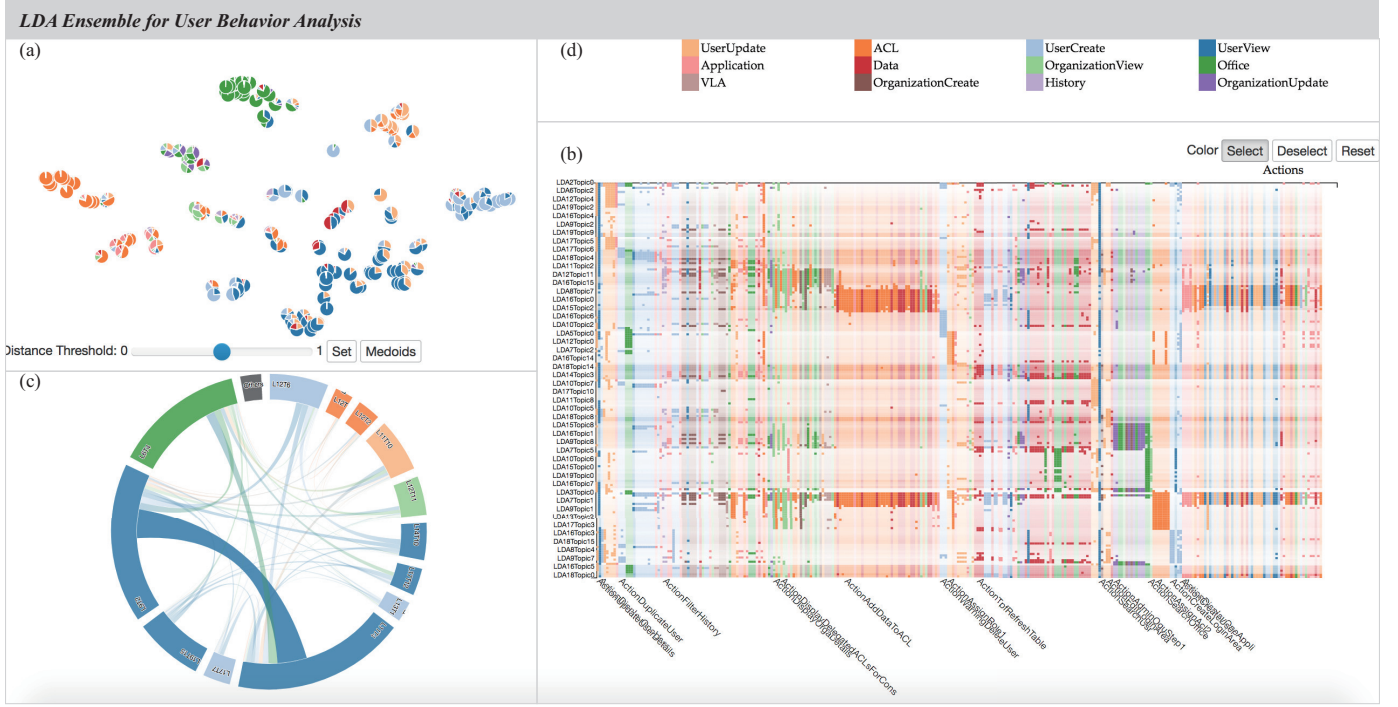


Fig. 1. Visual interface for exploring results of ensembles of topic models. The visualization system includes (a) topic projection view, (b) topic-action matrix, (c) topic chord diagram, and (d) color labels of action classes. We label the topics as “LDA i Topic j ” to reflect the LDA run and topic number.

using other criteria such as distinctiveness and coverage. *Distinctiveness* means clear differentiation between the types of behaviors represented by the topics. *Coverage* indicates the proportion of behavior instances that can be affiliated to some behavior types with a high degree of certainty. When the behaviors lacking a clear affiliation with any of the behavior types are few, they can be treated as atypical or possibly anomalous and thus may require a special attention from the analyst. We propose interactive visual tools, namely, a chord diagram and a matrix display (Fig. 1), for visual exploration and comparison of selected topics and assessment of the distinctiveness and coverage of a selected subset of representative topics.

Our contribution can be summarized as follows:

- We investigate and demonstrate the suitability of topic modeling (as a general class of techniques) for analysis of activity logs with the aim of identifying typical behaviors
- We develop an analytical workflow for behavior analysis and categorization, combining (i) generation of ensembles of models using topic modeling methods and (ii) interactive visual exploration of these models and selection of a subset of topics representing typical behaviors
- We develop a set of visual analytics tools to serve as an example of how the proposed workflow can be supported.

The paper is structured as follows. Section 2 explains the goal and scope of our study and provides the rationale for the choices we made regarding the approaches and methods. Section 3 reviews the related work. After describing the data and introducing the LDA ensemble method in Section 4, we present the visual analytics workflow for identifying behavior categories in Section 5. We demonstrate

the use of the workflow in two case studies (Section 6). Finally, we discuss the limitations and future work (Section 7).

2 RESEARCH SCOPE

This section provides justifications for the core decisions made to define the subject and course of our study.

Why use topic modeling for behavior analysis? Topic modeling is a well-established and widely used class of statistical techniques for obtaining summarized representations of semantics of text documents based on combinations of words occurring in them [3]. There is a similarity between the task of semantic characterization of text contents and the task of identifying behavior types in action logs. In text writing, the intended topic motivates the choice of words that are used to convey it; hence, the topic can be reconstructed based on the chosen combination of words. Similarly, an actor’s intent motivates the choice of actions that are used to fulfill this intent; hence, the latter can be reconstructed based on the chosen combination of actions. The chosen actions may be indicative not only of the intent but also of the actor’s strategy and preferences; therefore, we use the term “**behavior**” to refer to a conjunction of factors that motivated an actor to do particular actions.

Topic modeling methods are applied to large collections of documents for identifying common topics that occur in multiple documents. We conjecture that these methods may also be helpful for identifying common types of actors’ behaviors that are realized in multiple action sequences. This is what we test in our paper.

Techniques that are typically used in analyzing action logs pay primary attention to the sequence in which actions appear. There exist many sequence mining algorithms [4], [5] as well as systems for visual analysis of action or event

sequences [6], [7]. In topic modeling, texts are represented as “bags of words”, ignoring the sequence in which the words appear. Hence, topic modeling can by no means substitute sequence analysis methods, and this is not what we aim at. What we try to achieve by using topic modeling is capturing the semantics of action logs for understanding actors’ behaviors. From this perspective, which particular actions were taken is more important than the order in which they were performed. Again, there is an analogy with texts: whereas word ordering is important in building sentences and ordering of sentences is important for presenting ideas and arguments, the overall content of a document can be sensibly represented by a combination of key words or terms regardless of their ordering in the document.

Topic modeling versus clustering. We also considered possible use of clustering as another approach to extracting action log semantics. The approach would involve clustering of action logs by similarity, e.g., based on the occurrence frequencies of different actions or their TF-IDF scores [8]. The resulting clusters could be represented by vectors composed of statistical summaries of the original attributes, such as the means or medians of the values. This is similar to a result of topic modeling, which consists of vectors of action probabilities. There are several reasons for preferring topic modeling over clustering. First, an acknowledged challenge for clustering is the “curse of dimensionality” [9]. Clustering techniques depend critically on the measure of similarity, or distance, between data points. With the increasing of the number of dimensions, the data become sparse, and the distances between points tend to become relatively uniform, which makes division into clusters quite arbitrary and unstable regarding slight changes of parameters. Unlike clustering, topic modeling techniques are designed to deal with high-dimensional and sparse data, being based on word co-occurrences rather than distances between documents.

While the problem of high dimensionality can be alleviated by applying dimensionality reduction methods, the reliance of clustering on distances makes this approach not quite suitable when many action logs may contain instances of two or more behavior types. Clustering will not put such action logs together with other action logs where the same behavior types occur alone or in other combinations due to high distances to those action logs. Without a limitation on the required number of clusters, a clustering algorithm would put each combination of behavior types in a separate cluster. Additional efforts would be required to find commonalities between clusters. As the number of clusters is usually limited, diverse combinations of behavior types may be put in the same cluster, which will be very hard to interpret. In contrast, a basic premise of topic modeling is that any document may concern several topics, possibly, in different proportions; i.e., each document is modeled as a topic mixture. Topics are identified based on term co-occurrences rather than inter-document distances. Therefore, even if there are no documents solely concerned with some topic A but there are documents where A is combined with various other topics B, C, D, a topic modeling algorithm is quite likely to find topic A, because the co-occurrence frequencies of the terms pertinent to A will be high while their co-occurrences with the terms specific to B, C, D will be relatively less frequent.

Why LDA? In our study, we wanted to investigate the general suitability of topic modeling as a class of techniques and even as a concept for identifying and understanding behavior types. In this regard, the choice of a specific algorithm is largely arbitrary. We chose LDA [10] because it is a highly popular method implemented in multiple open software libraries and known as more general than Latent Semantic Analysis (LSA) [11] and Probabilistic Semantic Analysis (pLSA) [12]. Another quite popular method is nonnegative matrix factorization (NMF) [13], [14]. There are no principal differences between LDA and NMF regarding the inputs and outputs; therefore, NMF could be used for the purpose of behavior analysis in the same way as LDA.

Why ensembles of topic models? Most topic modeling methods require the user to set the number of topics that need to be extracted. An appropriate topic number may be unknown and hard to guess in advance. Another problem is that many topic modeling algorithms are non-deterministic and produce slightly different results in each run even if the number of topics is the same [15]. Therefore, better understanding of the contents of a text corpus can be gained by examining and comparing results of multiple runs of topic modeling, i.e., an ensemble of topic models.

The comparison needs to be supported by a visual display showing similarities and differences among topics. This can be a projection plot obtained by means of some dimensionality reduction algorithm (MDS, t-SNE, etc.). In a projection, similar topics are located close to each other and dissimilar ones are distant from each other. A cluster of spatially close topic indicates the existence of an archetype topic that is discovered by multiple runs of a topic modeling method. In other words, it is the same topic represented slightly differently in different models. The fact that it was found more than once indicates its significance and trustworthiness. Topics that are scattered over a wide area in the plot may be computational artifacts rather than representations of really existing topics. If there are a few topics that are distant from others, the analyst can check their trustworthiness through additional runs of the algorithm with incrementing the topic number in each run. An occurrence of similar topics in the additionally obtained topic models is an evidence of topic significance. Topics that remain far from all others can be ignored. In this iterative and interactive way, the analyst can gain understanding of what and how many topics exist.

Of course, this approach is not scalable to the number of topics and thus cannot be applied to a heterogeneous corpus of documents where the number of topics is very large. However, it can work very well on action logs where the vocabulary (i.e., set of possible actions) is very limited compared to the vocabularies of natural languages.

It may be argued that a suitable number of topics can be determined automatically based on some quality measures [16], [17]. However, what is optimal in terms of statistical indicators is not necessarily the most meaningful and useful to a human. An analyst may gain better understanding and knowledge of existing behavior types from seeing and exploring a topic space with dense and sparse areas, clusters and outliers.

Why not hierarchical topic modeling? It may seem that the use of hierarchical topic modeling methods [18]

removes the need to specify the number of topics in advance and can thus be superior to using ensembles of flat topic models. However, hierarchical topic models have their problems [19], [20], [21]. One problem is that they tend to generate large hierarchies where the overall number of topics is very high. Exploring and understanding such a hierarchy may be challenging. Another problem is that very similar topics may occur in different branches of a topic tree. It is hard to spot these similarities and understand the differences between the topic semantics. It is also questionable whether the idea of topic hierarchy (from more general to more specific topics) is transferable from linguistics to behavior analysis, i.e., whether it is valid to assume the existence of a hierarchy of behavior types. This can make sense in some applications but, possibly, not everywhere, and it may be quite hard to judge whether a computationally constructed hierarchy reflects real relationships between behavior types.

Character of the research results. This paper presents an approach that we developed in the course of the study. We describe a set of techniques that can support the approach; however, this is not a fully operational software system oriented to end users. In our study, we evaluated the basic idea and the approach to implementing it but not the utilization of particular software tools.

3 RELATED WORK

We discuss works in topic modeling, topic visualization, interactive topic editing, and visual analysis of behaviors.

3.1 Topic modeling

Topic modeling approaches emerged from the information retrieval theory [22]. A predecessor is Latent Semantic Analysis (LSA) [11], which groups documents based on occurrences of common words and simultaneously groups words based on their co-occurrence in same documents. Relationships between words can also be analyzed using word embedding methods [23], which represent words by multidimensional vectors such that the distances between the vectors indicate the word relatedness. LSA was further developed into Probabilistic Semantic Analysis (PLSA) [12], establishing a statistical probabilistic formulation of the technique. Latent Dirichlet Allocation technique (LDA) [10] is a formalism that is closely related to PLSA. The basic idea of LDA is that documents are represented as mixtures of latent topics, where each topic is a distribution over words of a chosen base vocabulary [10]. The distribution of topics, documents and words is modeled using Dirichlet priors (distributions). The PLSA technique can also be modeled as a LDA technique by using special Dirichlet priors. LDA is therefore more general than PLSA. LDA was successfully applied to many different domains, from movement to consumer behavior characterization [24], [25], [26]. In this work, the choice of a specific topic modeling method (LDA or other) is not essential, as explained in Section 2.

There exist many variants of LDA techniques. For example, Labeled LDA [27] was introduced to find most appropriate tags for text documents among all tags assigned by numerous readers. This gave rise to the idea that topics

might be used as a basis for improving the classification of documents [28]. Dynamic Topic Modeling [29] was proposed for applying LDA to time slices of evolving data. There were attempts to integrate human knowledge in the process of topic modeling, for example [30], [31], which can be regarded as human-in-the-loop topic modeling. All of them exploit the ability of humans to judge the correspondence of words and documents to particular topics on the basis of textual information. Generally, involvement of a human in the topic modeling process is important for choosing appropriate parameters, understanding the generated topics, and creating a good set of topics for summarizing the content of given documents. These ideas motivate us to develop an interactive visual analytics process which enables a human analyst to understand, explore, and select good candidates from ensembles of topic models without a prior assumptions of optimal parameter settings.

There are topic modeling algorithms that automate the selection of the topic number, which include hierarchical topic models [18], LDA using KD-divergence measures [16], GSDMM [17], etc. These methods propose their own criteria for determining the suitable number of topics, such as perplexity [18] and divergence measure [16]. All approaches have their advantages and limitations, and there is no single approach suitable for all situations. In Section 2, we provide arguments for enabling a human analyst to explore multiple topic models and select a subset of topics that are the most meaningful and useful from the analyst's perspective.

3.2 Topic visualization

Topic modeling is a complex process, in which visualization helps the analyst to interpret, understand and explore the process output, i.e., the relationships between documents, topics and words. Representation of a document as a mixture of topics, where a topic is a probability distribution over words, makes a *matrix* a natural representation to visualize the document–topic and topic–word distributions [32], [33]. Interactive matrix visualizations support reordering of rows and/or columns to reveal their hidden relationships. In our work, a matrix is used to visualize the topic–action distribution (Section 5.2) with the topics ordered by similarity (discussed in Section 5.1) and actions ordered by their probability values for the topics. Such ordering method could help users to understand which topics are similar and why they are similar with regard to the actions involved.

Alternatively, a *node-link diagram* can be used for the same purpose. Rönnqvist et al. [34] map the topic–word distribution to a force-directed graph layout with nodes representing both topics and words. Links connect words and topics with distances proportional to the corresponding probability values. ContextTour [35] superimposes a *contour map* over a word graph to depict the density distribution of all words. FacetAtlas [36] adds colored curves on top of the map to display different types of relationships between words. For exploring relationships between topics, we project them into a 2D space based on the word probabilities (Section 5.1) and build chord diagrams showing the topic similarity in terms of shared documents.

Documents may be associated with timestamps. Tiara [37] and Visual Backchannel [38] display evolving

topics using the river metaphor [39], mapping the number of documents concerning a topic to the width of the topic flow at every time segment. Beyond changes in topic magnitudes, TextFlow [40] explores relationships between topics over time, focusing on how the topics appear, disappear, split and merge. When the number of topics is large, it is challenging to understand them and their relationships. This can be facilitated by organizing the topics into a hierarchy and supporting its exploration [41], [42]. TopicPanorama [43] supports exploration and comparison of topics discussed in texts from multiple sources. Our support to exploring a large number of topics is a projection display showing topic similarities and interactive tools for selecting and examining representative topics.

Beyond examining topics from a single model, topics from multiple models have also been compared. Alexander and Gleicher [44] compute topic similarity using the word vector representation of a topic and align topics in two different models to show how they match or mismatch. The authors also use the document-topic matrix to cluster documents, allowing them to compare two topic models on how they affect the clustering results. Chuang et al. [45] propose a metric for assessing the topic quality by matching topics against reference concepts defined by domain experts. This approach is hardly generalizable to contexts like ours, where there are no predefined concepts that could be used as ground truth. Another work from Chuang et al. [15] examines the stability of topics from multiple algorithm runs with identical parameters. That paper echoes the motivation of our work regarding the unstable nature of topics produced by probabilistic modeling algorithms. Our paper proposes an ensemble method that combines results from multiple modeling runs and supports comparisons and interactive selection of suitable topics.

3.3 Interactive topic editing

The complex, often stochastic, and parameter-dependent nature of topic models motivated the development of interactive methods that enable analysts to intervene in the topic modeling process. Lee et al. [46] utilize results of topic modeling to identify “soft” clusters of documents. Their approach is one of the earliest examples of providing interactive ways to refine topic models by adjusting the keywords, filtering the data, and directly manipulating the clusters through operations such as splitting, merging or removing. Hoque and Carenini [47] note that results of topic modeling can be noisy and fail to match users’ information needs. They propose a topic modeling system for asynchronous conversations that revises the model on the fly on the basis of users’ feedback. Pavia et al. [48] devised a method where analysts provide correct labels for misclassified data instances (similar to *active learning* [49]) to improve classification performance. Choo et al. [50] comment on the inconsistency of topic models when they are run multiple times. As a solution to this, they suggest the use of non-negative matrix factorization as a more stable approach and support interactive editing of a topic model through keyword refinement, topic splitting, merging, and creation. Kim et al. [51] present an approach that can be called progressive topic modeling, in which a selected topic

can be refined by applying the topic modeling algorithm to the subset of documents affiliated to this topic.

The specifics of our problem renders the approaches involving interactive improvement of a topic model unsuitable. The analyst may have no preliminary ideas as to what kinds of topics (i.e., behavior types) might exist, and topic goodness cannot be judged just from seeing the most representative action labels, which is different from seeing combinations of well-known words of a natural language.

Alexander and Gleicher [44] extend the set of core tasks in topic modeling from building a single model to comparing multiple models. The authors propose methods to align and compare topic models computationally, and resort to visualization techniques to communicate the computed similarities and differences. El-Assady et al. [52] introduce a procedure of progressive learning of topic models and their parameters reinforced by analyst’s feedback from topic comparison.

The potential of involving human analysts in topic modeling has also been investigated in the machine learning literature. Andrzejewski et al. [53] propose an approach in which domain knowledge is incorporated into topic modeling by specifying “must-link” and “cannot-link” relationships between terms, i.e., what terms are likely or unlikely to have high probabilities in the same topics. Hu et al. [54] suggest an enhanced tree-based topic model that can incorporate analysts’ feedback on the quality of topics by re-learning the aspects that analysts find inappropriately represented. The authors argue that such interactive methods lead to richer and better adopted models and discuss the important role of involving human experts into machine learning approaches.

In our work, we build on these studies and draw ideas and inspiration from them. In most of the approaches discussed, the emphasis is on refining a single topic model or on comparison of two models. Our approach differs by generating multiple topic models, which add a new layer of information for analysts to interact with.

3.4 Visual analysis of behaviors

Behavior is a broad concept being studied in diverse disciplines such as social media analysis [2], mobility [55] and security [6], [56]. Representative tasks are pattern extraction [57], relationship identification [58], and categorization of individuals based on their behaviors [4]. Xiong et al. [59] used a flower metaphor for a graphical representation of message posting activities of people. Kanda et al. [58] analyzed movements of museum visitors searching for different visiting patterns. Orellana et al. [60] studied relationships among players of a mobile game. Personal behaviors may be very complex and often represented by heterogeneous data [61], [62]. Blascheck et al. [62] combine multiple different types and sources of data for analyzing behaviors of participants in user studies. Koldijk et al. [61] analyze multi-faceted behaviors of people at work. Rattenbury and Canny [14] apply a variant of topic modeling to action logs of computer users to discover so called “context structures”, i.e., combinations of files, folders, web pages, applications, and e-mail addresses that are used at the same or close times and, therefore, are likely to be related. The results are

presented to the users as visual summaries of their work, allowing them to reflect on how their day-to-day work routines are organized.

User behavior analysis has also been used for identifying anomalies [63]. Pannell and Ashman [64] demonstrate the potential of using such approaches in the context of cyber security through building user profiles based on features such as keystroke profiles and applications used. Even though this area is growing, the use of visualization to support behavior analysis has not been discussed yet.

Differently from the previous works, our research studies the suitability of the topic modeling concept for identifying latent behavior types and defines the visual analytics methods and procedures needed for supporting this task.

4 LDA ENSEMBLES FOR BEHAVIOR DATA

In the following, we describe our approach by defining our core terminology and explaining how topic modeling is used. We define a set of domain agnostic analysis tasks for behavior categorization through topic modeling. We illustrate the concepts with an example based on data stemming from a Logon and Security Server (LSS) management and control system as gathered in the context of a multi-disciplinary research project.

4.1 Definitions and data description

We define a *behavior*, or *behavior instance*, as a combination of *actions* performed by some actor during a period of time. One actor may have one or multiple behavior instances. Actions may belong to specific classes.

Our illustrative example, LSS, is based on logs from a digital application where the user base of an organization is managed through user authentication, access control and sophisticated user rights assigned to individuals and offices. The LSS data set consists of 14,929 sessions performed by system operators; A session consists of multiple actions, for example, “Search User”, “Display One User”, “Create User”, “Create Login Area”, etc. There are 296 distinct actions. These actions can be grouped into classes by domain experts. For example, the actions for searching and displaying a user can be in “User View” class, while creating a user or a login area can be in “User Create” class.

The existence of action classes is not essential for the idea of behavior modeling with LDA. However, it is difficult to represent a large number of individual actions in visual displays in a comprehensible way. Similarly to Monroe et al. [5], we utilize action classes for creating simpler views. Another example of using action classes can be given based on the IEEE VAST Challenge 2015 story [65], in which visitors of an amusement park attended attractions grouped in 12 classes, such as “thrill rides”, “shopping”, “food”, etc. Here, the actions are the visits of different attractions, and the attraction classes can be treated as action classes.

We support the use of several alternative classifications when available, as this can help to (a) vary the level of detail (bigger/smaller sets of action classes, different levels in a class hierarchy, if it exists) and (b) look at modeling results from different perspectives (different themes).

4.2 LDA for behavior data

The LDA technique was originally proposed by Blei [10] to detect latent topics in a collection of **documents**, each of which contains a set of **words**. The input to LDA consists of a document-word matrix DW , where DW_{ij} represents the frequency of word j in document i . LDA produces a set of topics T , a topic-word matrix TW , and a document-topic matrix DT . TW_{kj} specifies the probability of word j in topic k , while DT_{ik} represents the probability of topic k in document i . The probabilities are numeric values ranging between 0 and 1.

In our approach, we treat the **actions** as **words** and each **behavior** containing multiple actions as a **document**. Thus, we can find latent **topics** of behaviors (Fig. 2), which can be seen as distinct behavior types, or categories.

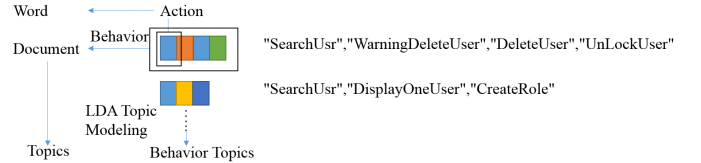


Fig. 2. The idea of using LDA for behavior analysis: analogously to deriving topics from documents consisting of words, we can derive topics representing behavior categories in terms of actions.

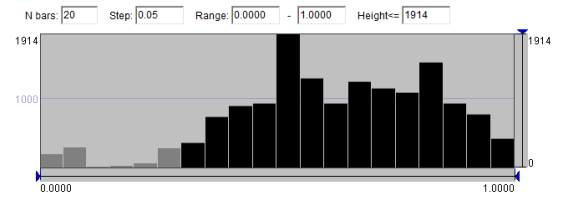


Fig. 3. The use of a histogram of the distribution of the maximal topic probabilities for choosing a suitable value for the behavior-topic association threshold. The black bars correspond to the probabilities 0.3 or higher.

We shall use the terms **affiliated** and **associated** for denoting relationships between behaviors and topics. A behavior is **affiliated** to a topic if this topic has the largest probability for this behavior. A behavior is **associated** to a topic if its probability of this topic is larger than a given threshold; hence, one behavior can be associated to zero, one, or multiple topics. A meaningful threshold value can be chosen based on inspecting the shape of the statistical distribution of the maximal topic probabilities for all behaviors, so that only a small fraction of behaviors remains not associated to any topic while the threshold value corresponds to a notable upward turn in the shape of the probability curve. At the same time, the threshold value should not be too low, to avoid many behaviors being associated with two or more topics. In our case, we found 0.3 to be a good threshold value; see Fig. 3. With this value, 852 behaviors out of 14,929 (i.e., 5.7%) are not associated to any topic. Such behaviors are put in a special group called “other”.

The LDA method has parameters α and β controlling the distribution of topics over documents and words over topics. We follow established guidance on how to identify appropriate values for these parameters [10], [66] and choose $\alpha = 0.1$ and $\beta = 0.1$ for the following cases.

Apart from α and β , LDA requires setting of parameter N , which is the number of topics to extract. An appropriate topic number is often not known in advance. To cope with this problem, we propose to use LDA ensembles, where LDA runs multiple times with different values of N . The combined results of such an ensemble provide a set of candidate topics to select from.

4.3 Tasks in behavior categorization with LDA

Due to the lack of explicit semantics associated with behaviors as discussed earlier, the interpretation of LDA results obtained for behaviors is more challenging than for textual data. For example, the relationship between actions “search office” and “assign role” is not as clear as the semantic relatedness between words “Germany” and “European”. Hence, we need to provide tools helping analysts to understand, compare, and evaluate behavior topics and on this basis select appropriate topics representing behavior categories. The tools need to support the following tasks:

- **T1: See and understand similarities and differences among topics.** Groups of similar topics coming from different LDA runs are highly likely to represent inherently existent behavior types.
- **T2: Explore the features of the topics.** The features characterizing the topics include the action probabilities for the topics, the sets of affiliated and associated behaviors, and the action frequencies in these sets. By analyzing these features, analysts can gain a deeper understanding of the topic properties.
- **T3: Make a selection of representative topics.** The analyst must be able to make informed selection of representative topics and interactively modify the selection made.
- **T4: Evaluate the topic selection.** The analyst should be able to check the suitability of a selected topic subset for categorization of behaviors. The topic subset is evaluated regarding the distinctiveness among the topics (i.e., low intersections among the topic-associated groups of behaviors) and the coverage of the behavior set (i.e., low number of behaviors that are not associated with any topic).

5 INTERACTIVE VISUAL ANALYTICS OF LDA ENSEMBLES

To support the defined tasks, we propose a suite of tools (Fig. 1) including topic projection, matrix visualization of topic features, chord diagram for groups of topic-associated behaviors, and multi-level color labels of actions (Fig. 4).

5.1 Projecting multiple ensembles of topics

To reveal similarities and differences among topics (T1), commonly used projection (a.k.a. embedding) methods, such as MDS [67] or t-SNE [68], can be employed for projecting the topics onto a 2D space (Fig. 1a). Such methods proved to be suitable for visualization of topic modeling results [50], [51], [69], [70]. The projection is based on the inter-topic distances derived from the action-in-topic probabilities. The distance between two topics is determined by the

accumulated differences of the probabilities of action occurrences in the topics (accumulated Manhattan distance [71]). In the visual display of the projection results, topics are represented by pie chart glyphs showing the probability distributions of action classes. The segments in each pie chart are drawn in the order of decreasing probabilities of the action classes. The analyst can apply different action classifications to vary the level of detail and/or the themes. From the glyph distribution in the projection space, analysts can assess the similarities and differences among topics and on this basis interactively select representative topics. The visual similarities and differences between the glyphs themselves provide additional information for judging topic similarity, so that the analyst does not have to fully rely on the distances in the projection. The commonly acknowledged disadvantages of pie charts (see, e.g., [72]) refer mostly to the accuracy in estimating and comparing numeric values. In our case, the pie charts are intended to be used not for estimating exact values and differences but for noting whether closely located topics appear sufficiently similar. No detailed examination of the pie segments is required; the pies can be perceived as wholes; moreover, clusters of close similar pies can also be perceived holistically (Fig. 1a). The combinations of the colors occurring in different parts of the projection display give a general idea concerning the respective kinds of behaviors.

The projection display serves as an overview, an interactive panel, and a representation of the current selection.

5.2 Visualizing features of topics

To gain a detailed understanding of the topic features in terms of the action probabilities (T2), we visualize the data from the topic-action matrices TW (Section 4.2) received from the LDA ensemble by means of a matrix display (Fig. 1b). The columns correspond to the actions and the rows to the topics (all or selected). The colors in the cells encode the action classes. The vertical ordering is based on the similarity of the topics, which is consistent with the topic projection view. The horizontal ordering is done according to the descending action probabilities following the vertical arrangement of the topics. This means that actions with high probabilities for the first topic are put in the beginning, following by not yet included actions with high probabilities for the second topic, and so on. Such an ordering method is intended to facilitate observing the commonalities and differences between similar topics.

To support the identification of the key actions involved in the topics, we encode the action probability by the color opacity. The linear mapping of the probabilities to the opacity levels may not work well for probability distributions with high peak values. To mitigate this problem, we propose a mapping involving a threshold t .

$$Opacity(x) = \begin{cases} 0.9 * (x - min)/(t - min), & x < t \\ 0.1 * (x - t)/(max - t) + 0.9, & t \leq x \end{cases} \quad (1)$$

Depending on the chosen threshold, different information may be conveyed by the matrix display. Fig. 5a corresponds to $t = max/10$. The display highlights dominating actions in the topics but loses less pronounced topic features. This representation is suitable for determining the most representative actions in the topics. To avoid losing potentially

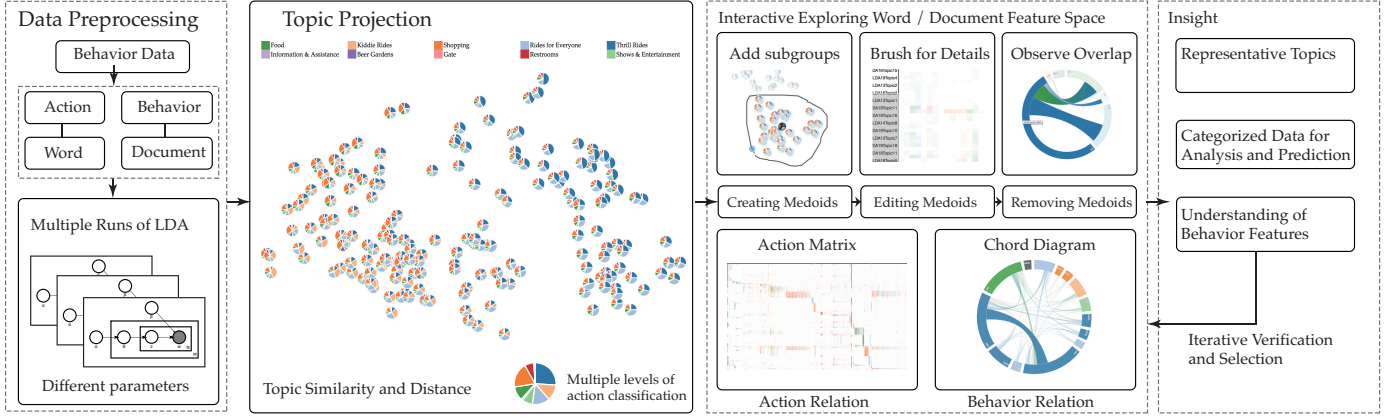


Fig. 4. Visual analytics tools enabling interactive extraction and exploration of behavior categories with the use of topic modeling. Analysts can select representative topics and modify the selection after understanding the topic features, their capabilities to differentiate behaviors, and the coverage of the behavior set.

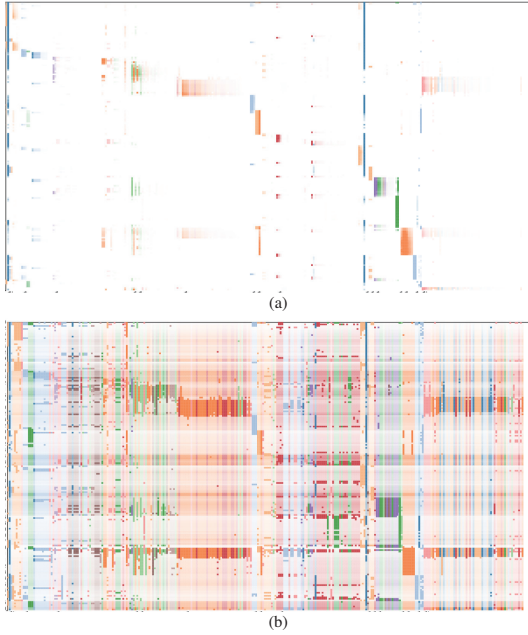


Fig. 5. Matrix color opacity scale settings. (a) Highlighting extreme values. (b) Showing details of value distributions.

important details, the threshold can be set based on the outlier definition from Tukey's fences [73]. The threshold is calculated based on the interquartile range.

$$t = Q_3 + k * (Q_3 - Q_1) \quad (2)$$

In Fig. 5b, k has been set to 3. One can observe both dominant actions and actions with lower probabilities.

The representativeness of a current topic selection with respect to the behavior set can be judged based on the amount of intersection between the topics, the numbers of the associated behaviors, and the number of behaviors without topic association. To support judgments based on the topic intersections, we propose to use a chord diagram (Fig. 1 c), in which the selected topics are arranged in a circular layout and connected by arcs that express their relationships. The width of an arc represents the amount of intersection between the connected topics, i.e., the number of behaviors associated to both topics. The topics in the diagram are shown in the colors of their dominant actions,

and the same colors are used for the arcs. When an arc connects topics with different colors, it is divided into two parts having the same colors as the topics at the respective arc ends. In response to interactive hovering on an arc representing a topic, the links of this topic and the linked arcs are highlighted and the remaining arcs and links faded. This facilitates observing intersection relationships of any particular topic.

To provide more details for the behaviors, we propose a matrix view of affiliated behaviors, which is similar to the previously described matrix of topics, but here the rows correspond to behaviors rather than topics. Both matrices have consistent ordering of rows and columns. The behavior matrix shows features of the topic-affiliated behaviors in terms of the action frequencies (Fig. 9). High similarity between the topic and behavior matrices indicates good representativeness of the topic selection. Inconsistencies, such as existence of actions with high frequencies in multiple behavior groups while their probabilities in the respective topics are low, indicate the need of updating the selection by adding topics with high probabilities of these actions.

5.3 Selection of representative topics

The proposed approach to topic selection (T3) is based on interactive grouping of topics by similarity and picking a representative topic from each group. The analyst defines a topic group by interacting in the projection display. The group medoid, i.e., the topic with the minimal distances to all others, is automatically selected as the representative. Currently selected medoids are highlighted in all views by black strokes (Fig. 1). Alternatively, a representative topic could be created by computing the average or median probability values among all candidates. The medoid approach has two advantages. First, the medoid is not affected by extreme values in outliers, in contrast to the average value. Second, the medoid is one of the existing topics with known probabilities for the behaviors whereas constructing a new representative topic requires additional calculation of its probabilities.

5.3.1 Interactive definition of topic groups

We support two alternative operations for creating groups. First, the analyst can define a group of topics by using a

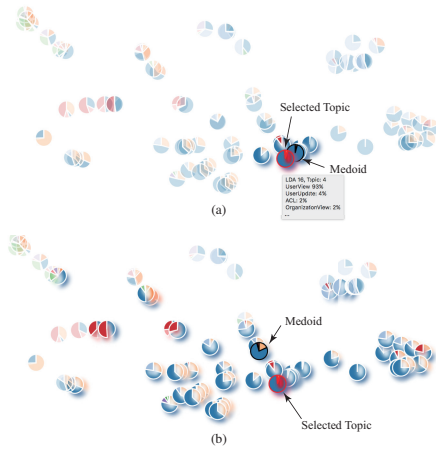


Fig. 6. When the analyst clicks on a topic of interest, the system selects other topics within the distance less than a given threshold: (a) using a small threshold (0.1), (b) using a large threshold (0.7).

free-form brush for outlining a visually perceived dense cluster of topics in the projection display. In this operation, the analyst considers not only the relative positions of the topics in the projection space but also the visual similarity of the corresponding pie charts. Taking into account that the distances in the projection space may not be proportional to the distances in the high-dimensional feature space, we also enable the second interaction method. The analyst can click on a topic of interest, and the system responds by selecting all topics within a given distance from this topic in the feature space (Fig. 6). The medoid of the group defined in this way may not be the same as the originally selected topic.

5.3.2 Group exploration and editing

Multiple interaction techniques support the exploration of topic groups (Fig. 7). The action probabilities for selected topics are visualized in the matrix view. For example, Fig. 7c exhibits several subgroups with different combinations of dominating actions. The actions mostly belong to two classes, ACL (Access Control List) and Applications. The class ACL (orange color) includes “Add data to ACL, Create ACL for role, Filter roles for ACL, Delegate ACL”, etc. The class Application (red color) includes “Display one application, Manage app role, Filter roles for app, Find app in tree”, etc. The columns and rows can be filtered by brushing the axes in either select or unselect mode. Thereby, the analyst can select regions of interest and drill down to a detailed view of the action-based features of the selected topics (Fig. 7d). In this example, the analyst has brushed the group G1 on the Y-axis and deselected the actions with low probabilities by brushing the X-axis. The detailed inspection may suggest a need in subdividing the group into subgroups (G2 and G3), as in Fig. 7b. Group subdivision is done by interacting with the matrix, and the results are reflected in the projection view.

After exploring a group of topics, the analyst may want to expand or shrink it. Groups are edited by direct interaction in the topic projection. For example, analysts can re-brush on the topic projection to update the groups (Fig. 7a).

The analyst can also dynamically change the threshold that is used for defining groups based on the topic distances to representative topics. It is done by dragging the slider in the control below the topic projection view (Fig. 1a). In the chord diagram, the analyst can observe whether there are large intersections between selected topics, and, if needed, modify the corresponding groups in the topic projection.

5.3.3 Removing a topic from the current selection

The analyst may decide to remove a topic from the current selection if the number of associated behaviors, which is represented in the chord diagram, is too small. Another possible reason for removing a topic from the selection is a high intersection with another selected topic. The analyst can remove a topic by clicking on the topic band in the chord diagram. Since each topic in the selection is a group representative (medoid), topic removal results in destroying the group it represents, i.e., the topics become ungrouped. The analyst can then perform another selection among the unselected topics, to examine a new medoid.

5.4 Progressive categorization of behaviors

We formulate criteria for evaluating a topic selection (T4) and then introduce interactive operations that support the process of progressive definition of behavior categories through iterative assessment and revision of topic selection.

5.4.1 Criteria for evaluating topic selection

The selection of a subset of topics to represent behavior types is guided by the following criteria, whose relative importance may vary depending on the data properties and analysis goals:

- **Distinctiveness.** Understanding intersections between topics is needed for deciding if each topic is unique and irreplaceable. This criterion is discussed in more detail below the list.
- **Coverage of behaviors.** The selected topics should jointly cover most of the existing behaviors. The amounts of **associated** behaviors covered by the selected topics are shown in the chord diagram (Fig. 1c), and a special group labeled “other” consists of the currently uncovered behaviors.
- **Patterns of action probability distributions.** The distribution patterns of the actions, which are visible in the matrix view (Fig. 1b), indicate whether a topic is a general or a specific one.
- **Consistency between topic-affiliated behavior features and topics’ action features.** For a good topic selection, it is expected that the action frequencies in the affiliated behaviors are consistent with the action probabilities in the topics. The comparison of the behavior features to the topic features is supported by juxtaposed matrices (Fig. 9). When a mismatch occurs, the analyst can see which features are not correctly captured by the current topic selection.

The distinctiveness criterion is one of the most complex criteria. It would not be a right idea to just try to minimize the number of shared behaviors between topics because, as explained in Section 2, one particular behavior may be a mixture of two or more behavior types, and this is the

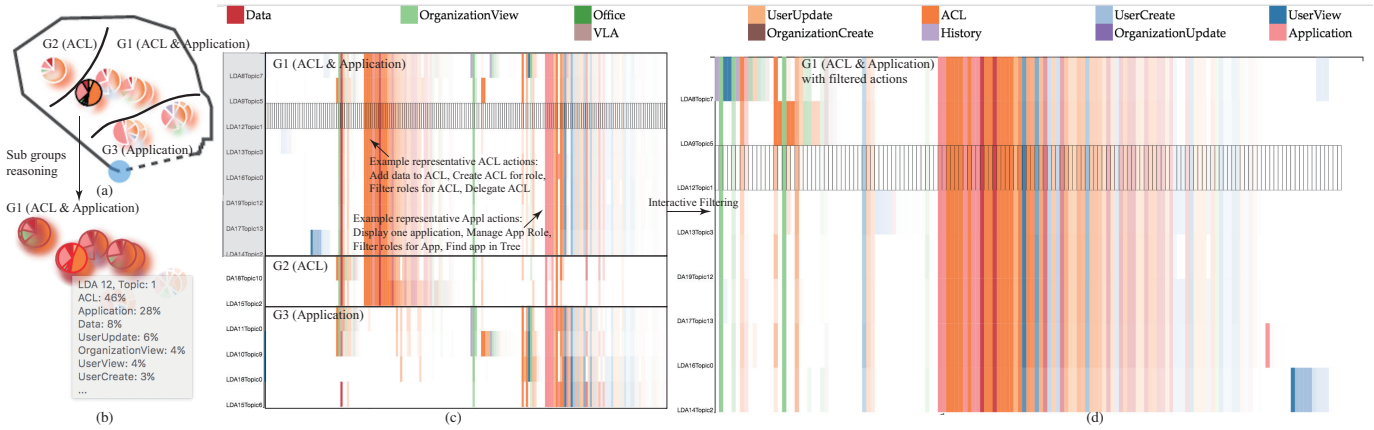


Fig. 7. Matrix exploration and editing of topic groups. (a) A topic group is selected in the topic projection, and (c) three sub-groups (G1, G2 and G3) are identified in the matrix view. By brushing in both axes for filtering, details for Group G1 are selected and shown in (d) matrix visualization and (b) topic projection.

main reason for using topic modeling rather than clustering. Hence, a high amount of shared behaviors between two topics observed in the chord diagram (Fig. 1c) should not be treated as a sign of alarm, but it indicates a need in a detailed inspection of these topics using the projection and matrix views (Fig. 1a,b). If it turns out that the topics are quite similar, it is reasonable to remove one of them or replace this pair of topics by one topic located in the projection view between them and thus likely to be similar to both these topics. If the topics are distinct but almost all behaviors associated with each of them are shared with the other, it is an indication of an excessive specificity of these topics and a high likelihood of the existence of a more generic topic that combines the features of both topics. The topic projection view can help the analyst to find such a unifying topic for replacing the pair of excessively specific topics. If the fraction of the shared behaviors is not very high, both topics can be considered suitable. Hence, the application of the distinctiveness criterion is highly reliant on the analyst's judgments based on interactive exploration of the topics.

5.4.2 Inspection and revision of topic selection

For selected topics, the analyst can explore the action probability distributions at different levels of detail by interacting in the matrix view. Action classes (Fig. 1d) can be selected or de-selected for considering actions of specific classes in more detail. Other views are updated according to the operation (Fig. 8).

Once the analyst has selected some topics, behaviors that are not associated to these topics (i.e., the probability is less than the threshold) are put into the special group "other". A good selection should make the "other" group as small as possible, but its size may not reach zero, since behavior outliers may exist. When the analyst clicks on the group "other" in the chord diagram, the yet not grouped topics to which the respective behaviors are associated are highlighted in the topic projection. This suggests the analyst where in the projection space the next group should be made. Blurred shading is applied to already explored topic groups (Fig. 6), informing the analyst of the current state of the process. These hints guide the analyst in exploring groups and adding new groups. Blurring is not applied



Fig. 8. By clicking on the color labels that correspond to action classes, the analyst can select or exclude specific classes of actions for/from the visualization. "User View", "User Create" and "User Update" classes are selected for investigating detailed distributions.

to group medoids, and their distribution in the projection space indicates how diverse they are.

The analyst may select representative topics for inspecting the action frequency distributions in the affiliated behaviors (Fig. 9). The matrix view of action frequencies shows how representative each topic is and whether not yet captured behavior patterns exist. For example, Fig. 9b reveals that there are combinations of actions from classes "User view" (blue) and "Office" (green) that are not represented by any of the selected topics. This indicates a need in adding a new topic. The analyst can find suitable topics with high probabilities of the underrepresented actions and create a new group, the representative (medoid) of which will be added to the current topic selection (Fig. 9c). The action frequencies view is updated reflecting the behaviors affiliated to the topic added (Fig. 9d).

5.5 Rationale for the visualization design choices

Here we explain how the visualization techniques were chosen in accord with the data specifics and analysis tasks.

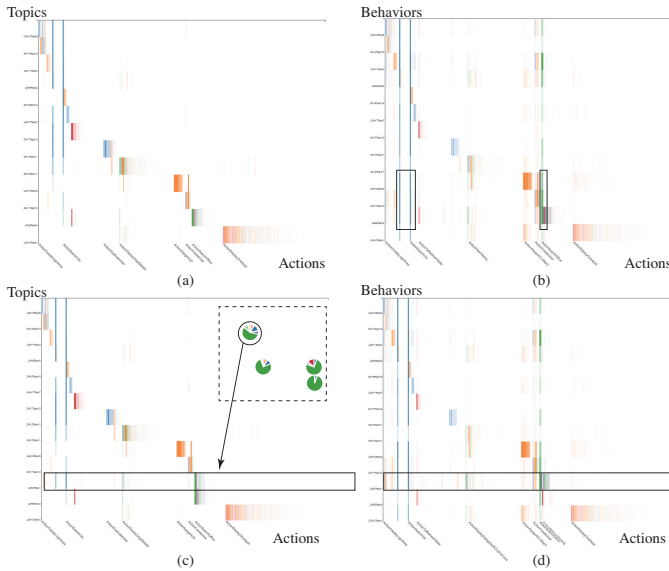


Fig. 9. Iterative exploration and revision of topic selection. (a) Action probabilities for the topics. (b) Action frequencies in the behaviors affiliated to the topics of (a). (c) For a group of poorly represented actions, the analyst adds a new topic to the selection. (d) The action frequency view is updated according to the modified topic selection in (c).

5.5.1 Visualization of inter-topics similarities

Visualization of inter-topic similarities is necessary for supporting task **T1** (Section 4.3). We do this using the technique of dimensionality reduction, which is a common approach to supporting detection of clusters (i.e., groups of similar items) in multidimensional data, in particular, LDA outputs [74]. This technique creates a two-dimensional arrangement of topics, where we use pie chart glyphs to represent the probabilities of action classes for the topics (Fig. 1a) in order to provide additional information for assessing the similarities among neighboring topics in the projection. The need in such additional information exists because the distances in the projection space may not accurately reflect the distances in the multidimensional space. Glyph components are ordered consistently for all topics, which facilitates comparisons. The use of the glyphs in the topic projection has been explained and discussed in Section 5.1. In addition, it can be noted that the common practice to represent topics by several most significant words [50], [51] is less suitable in our case, where the “words” are action labels. As we explained earlier (in Introduction and Section 4.3), semantic relatedness of actions may be hard to judge from their labels. Another problem is that action labels may be quite long, and putting them in the projection display may introduce much clutter. Instead, topic details, including the significant actions and their probabilities, are shown upon mouse hovering on the topic glyphs. Task **T3** (Section 4.3) is supported by interactive topic grouping operations in the projection view.

5.5.2 Visualization of action probabilities distributions

Visualization of action probabilities distribution is necessary for supporting task **T2** (Section 4.3). A matrix visualization is naturally suitable for this purpose since action probabilities for the topics exist in the form of topic-action matrices in LDA outputs. For consistency, the same visualization is used

to show the action frequencies in topic-affiliated behaviors (Fig. 9). Topic comparison and selection are supported by interactive operations.

5.5.3 Visualization of topic intersections

Detecting overlapping topics is one of the main subtasks in the evaluation task **T4** (Section 4.3). During the design process, we considered three candidate visualizations: Venn diagram, matrix, and chord diagram. A Venn diagram is good for showing set intersections when at most three sets overlap but does not scale to complex cases. Matrix visualization represents topics with rows and columns; hence, each intersection is shown twice, which is space-inefficient and can lead to confusion. Showing only half of the matrix removes the duplication of information but leaves space unused, and this issue is more noticeable with increasing number of selected topics, which also increases the size of the matrix. In contrast, a chord diagram has a constant size irrespective of the number of selected topics. In this display, the amounts of topic-associated behaviors are represented by the sizes of the arcs that make the outer circle, and topic intersections are shown by links inside the circle. A potential drawback is visual clutter due to link intersections. However, such cases indicate a need to improve topic selection; hence, the disadvantage transforms into an advantage for the analysis process.

6 CASE STUDIES

The main question we investigate here is: *Can topic modeling produce results that can be treated as representations of meaningful behavior types?* To answer this, we performed a case study in a particular application area (cyber security) and consulted domain experts to judge the results of the analysis. The answer was positive, which also means that the combination of tools and the procedure we used for performing the analysis worked well.

To test the generality of the approach, we conducted an additional experiment with synthetic data concerning customers’ behaviors provided for the IEEE VAST Challenge 2015 [65]. In both use cases, we used an ensemble of 18 LDA runs with the parameter N (topic number) ranging from 2 to 19, producing 189 candidate topics. Generally, the chosen upper limit for N should be slightly larger than the expected maximal number of distinct behavior types in a given domain.

6.1 User behaviors in LSS

As introduced in Section 4.1, this dataset contains 14,929 sessions (behaviors) with 296 distinct types of actions performed over a period of 30 days. These actions have been organized into 12 classes by domain experts. The number of performed actions in a single behavior ranges from 1 to 893, whereas a majority of the behaviors consist of 6 to 14 actions. The motivation for the analysis was to understand what behaviors are typical, with an outlook to use this information for detecting anomalies.

The domain experts are currently using a computational model that assigns anomaly scores from 0 (normal) to 1 (highly abnormal) to the behaviors. This current model handles actions in isolation, and thus, combinations of actions

are not considered. This eventually leads to inaccuracies and high numbers of false positives, leaving the experts in need of improving the model. One viable strategy is to investigate how the model performs for different behavior types in terms of action combinations. However, the behavior types are not known and need to be defined – the problem which our LDA-based approach is designed for. The analysis reported below was performed by the tool developers in cooperation with the domain experts, who provided their feedback on the results and set further questions that guided the analysis process.

6.1.1 Interactive exploration process

In a projection view of the results of the LDA ensemble run, we observe several dense groups of topics represented by similar glyphs (Fig. 1a). From these dense groups, we select representative topics, such as the topic “LDA6Topic4” with dominating actions from the class “Office” (Search Office, Display Office, etc.). However, not all topic groups are homogeneous, which is indicated by differences in the dominating colors. We brush the orange and pink topic regions (ACL and Application-related actions, respectively) (Fig. 7). By drilling down to the details, we find out that there are three distinct subgroups, G1 focusing on ACL and Application, G2 focusing on ACL, and G3 focusing on Application. Since the number of behaviors covered by the whole group is not large, we choose to select the medoid within G1 and thus avoid making several small behavior categories. The result represents the behaviors related to both ACL and Application actions.

In the regions in the projection space where the topics are not dense, we define groups by clicking on topics having more neighbors than others and using the function of automated selection of similar topics, for which we set a small distance threshold of 0.1 (Fig. 6a). As already considered groups are shaded by blurred colors, it is easy to see which parts of the projection space are not explored yet. We find that the remaining topics are mostly dominated by three classes of actions, “User View”, “User Create” and “User Update”. We focus on these action classes by unselecting the others for a detailed examination of the probabilities of the actions from these classes. This facilitates seeing the differences among the remaining topics in terms of the probabilities of the individual actions from these classes and helps us divide the topics into subgroups, whose representatives (medoids) are added to the topic selection. For example, seeing that the action “User Display” is often accompanied by “User Unlock”, we make a subgroup of topics where this pair dominates.

6.1.2 Analysis and modeling result

With the progressive exploration process, we come to a selection of 13 topics from different LDA runs, with the minimal number of 238 and maximal number of 3,445 affiliated behaviors per topic. The topics have been interactively examined and evaluated with regard to the criteria introduced in Section 5.4.1 using the corresponding tools. The selected topics are treated as representations of behavior types and make together a behavioral topic model. This model can be used for determining the types of existing and new behaviors in the following way. For each session,

the model determines the probability of each topic. If there is a single topic with a high probability while all others are low, this session is an instance of the respective behavior type. This means that the behavior is typical, i.e., normal. The highest topic probability value can thus be treated as a measure of behavior typicality, or normality. When all topic probabilities are low, it may mean an atypical, strange behavior, which may require detailed examination. When two or a few topics have relatively high probabilities compared to the others, the behavior may be a mixture of the corresponding behavior types. How normal this mixture is, can be judged from the amount of intersection between the topics, i.e., the number of shared associated behaviors in the historical data.

6.1.3 Representation of behaviors for expert evaluation

In the analysis, the relationships between the topics, actions, and behaviors were inspected using the matrix displays. However, this representation is not well suited for communicating analysis results to the domain experts. For them, it is more convenient to see only the essential (highly probable) actions of each topic and only the actions with sufficiently high frequencies in behavior groups, whereas the matrix displays include lots of irrelevant details. For experts’ convenience, we chose a different visual representation of behaviors. We created a two-dimensional projection of the set of actions based on the median pairwise distances between the action occurrences in the sessions. The actions that often occur together in a session tend to appear close in this “action space”. Each session can be represented as a trajectory in this action space. Multiple sessions can be aggregated into a flow map [75], in which actions are linked by arcs, as in Fig. 10, or by half-arrows, as in Fig. 11, with the widths proportional to the number of times when one action is immediately followed by the other. The probabilities or frequencies of the actions can be represented by sizes of dot symbols drawn at the actions’ locations. Actions with near-zero probabilities or frequencies can be hidden to reduce the display clutter and show only essential information. The flow map technique can be applied, in particular, to the sets of topic-affiliated behaviors. The resulting “maps” can serve as visual representations of the topics (Fig. 10). The flow map technique can also be used for representing a single session (Fig. 11) or a set of sessions performed by a user. The representation of behaviors and groups of behaviors as trajectories and flows in an artificial space leverages our previous experience on semantic representation of mobility behaviors [76].

6.1.4 Expert verification and feedback

For checking the validity of the behavioral topic model, the domain experts were asked to assess the model and its results and provide their feedback in respect to the following questions:

- 1) whether the selected topics can be treated as representations of sensible categories of normal behaviors of the LSS system operators;
- 2) whether high typicality values can be indicative of behaviors being normal;
- 3) whether low typicality values can be seen as an indication of a possible anomaly.

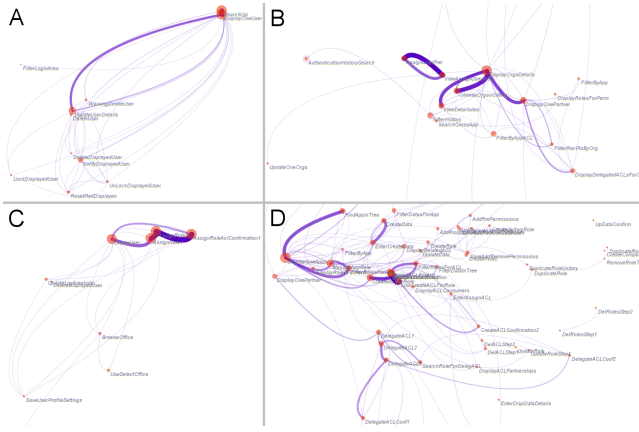


Fig. 10. Examples of topics represented by flow maps. A: Update user data; B: Check organization data; C: Create users and assign roles; D: Update application.

Validity of extracted behavior types

We assigned meaningful labels to the 13 topics selected based on the most typical actions in the topic-affiliated sessions. The labels are: *Create login area*, *Edit login area*, *Create users and assign roles*, *Remove a role of a user*, *Duplicate a user*, *Check user data*, *Update user data*, *Unlock user and reset password*, *Update application*, *Assign ACL*, and *Check organization data*. The topics were presented to the experts in the form of flow maps, as in Fig. 10. The experts agreed that 12 of the 13 topics can be seen as representing sensible behavior types. Selected examples of sessions affiliated to the topics were judged as normal behaviors. One exception was the topic “*Update application*”. In the expert’s opinion, this topic mixes application-focused and ACL-focused operations and should rather be replaced by two distinct topics. This suggestion was later fulfilled using the tools for editing topic groups (Section 5.3.2). Essentially, the experts’ feedback suggests the positive answer to the question 1.

Assessment of typical and atypical behavior instances

Examples of typical and atypical behaviors were presented to the experts in the form of flow maps. In general, the experts gave positive feedback noting that our approach to behavior modeling provided new investigating perspectives for them.

All examples of behaviors with high typicality values were judged as normal sessions by the experts, which gives a positive answer to question 2 and confirms the possibility to use the behavioral topic model for behavior type recognition. There were several examples with high typicality according to the topic model but also high anomaly scores assigned by the currently used model. The experts agreed that these behaviors are quite typical in terms of the actions performed and commented that the high anomaly scores could be false alarms of the current model or could be assigned based on other features of the sessions, such as the IP address or user information.

For obtaining an answer to question 3, the experts were shown examples of sessions with very low typicality values. These included sessions that received high and low anomaly scores from the current model. A few sessions are shown in Fig. 11. Out of 10 example sessions with high anomaly scores, 5 were judged as somewhat strange and others as rather normal despite of the high scores. Likewise, out of

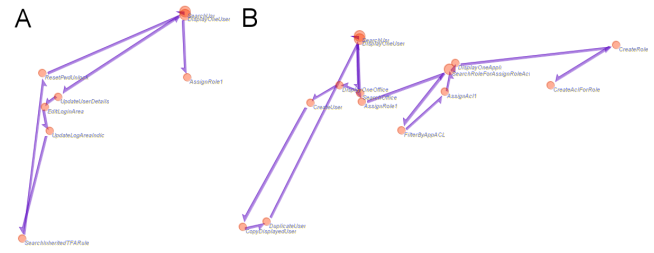


Fig. 11. Flow maps represent examples of sessions with high (A) and low (B) anomaly scores both having low typicality by the topic model.

8 examples with low anomaly scores, 4 were assessed as slightly abnormal. The experts noted that all sessions that received low typicality values from the behavioral topic model included actions that are specific for different LSS system’s functions dealing with users, offices, organizations, etc. Generally, such behaviors are not necessarily suspicious and indicating potential fraud. Still, behaviors mixing diverse functions may deserve a detailed examination.

The behavioral topic model also assigned low typicality scores to sessions that consisted of a single action repeated many times, e.g., “*Search User*”. These sessions had low anomaly scores by the current model. The experts acknowledged that, since such behaviors happened relatively rare, their current model might not be trained to treat them as abnormal. Hence, the new approach to session analysis made the experts aware of the existence of such behaviors.

User-level behavior analysis

The domain experts were highly interested in a possibility to categorize the LSS system users based on their behaviors in the sessions. The behavioral topic model allows characterizing the users based on the following criteria:

- whether a user has one or a few usual (repeatedly occurring) behavior types or has diverse behavior types;
- which specific behavior types are usual for a user;
- whether a user has occasional occurrences of behaviors of types that not usual for this user;
- whether a user has behavior instances with low typicality values and how frequent such instances are.

User profiles based on these criteria can be used for assessing new sessions with respect to their expectedness. The experts agreed that this may be a good approach to enhancing the security checks.

By considering the examples provided, the experts agreed that such behavior summaries can be a useful addition to the currently existing user profiles that are used in the security monitoring system. The users practicing one or two behavior types were judged as having routine behaviors. Regarding occasional occurrences of other behavior types among the sessions of such routine users, the experts said that such occurrences are not necessarily suspicious since they may correspond to functions that need to be performed only once in a while. Hence, the overall frequencies of behavior types need to be taken into account in deciding whether a user-unspecific activity may be suspicious.

Regarding the users whose sessions could not be associated with any behavior types, the experts were cautious in considering them as suspicious. They said: “*Generally, there is not enough data to take conclusion for users with only ten sessions. But for one user you provided there is a large amount of*

observed sessions, which is worth investigating”.

Feedback on Visual Analytics Process

In this work, we have a team including visual analytics experts and security experts who work together. Security experts have their needs and problems to solve. The topic modeling-based approach provides them with a different perspective to understand user behaviors. Since the main research question concerned the utility of topic modeling for behavior analysis, the visual analytics part of the team did not attempt to develop a polished software system that could be immediately used by the security practitioners. Therefore, the analysis was done by the tool developers under the supervision of the domain experts, and the latter were asked to comment on the analysis process and techniques used.

The experts’ feedback was positive. Concerning the LDA ensemble use, the experts stated that *“the choice behind LDA ensembles is clearly motivated, and the advantages of this method compared to the state of the art is also very clear”*. The domain experts have basic dashboard visualizations in their current working system. Compared to the dashboards, the interactive explorations that are possible in the visual analytics system are preferred by them, as well as the visualizations of behavior topics. They said: *“The interactive exploration capabilities of this tool for understanding behaviors are truly amazing. Thanks to these topics resulting from LDA ensembles and the way you visualize them interactively, we can have many ways to look at behavior data across so many dimensions. This can certainly help in the initial step of overall understanding user behaviors and categorizing them.”* They noted that the projection of topics with glyph design and lasso interactions are useful in the analysis process.

The experts especially appreciated the possibilities for grouping behaviors: *“The VA tool can support the discovery of new insights into complex behaviors, such as groups of frequent and/or similar behaviors as well as hidden associations among them. This can be really useful.”* The experts also pointed out several issues to improve, which we will discuss in the following discussion section.

According to a wish of the domain experts to have the visual analytics tools at their disposal to better understand the user behaviors, improve their original models, and detect anomalies more reliably, the tool developers started a process of integrating the system into the current work environment of the security specialists.

6.2 Visiting behavior exploration

In this case, our approach is applied to behaviors of customers of an amusement park. The data set originates from IEEE VAST Challenge 2015 [65]. The actions are visits of different attractions and other points of interest (POIs). There are 73 POIs grouped in 12 classes, such as “thrill rides”, “shopping”, “food”, etc. During 3 days, 11,374 visitors performed 1,562,421 actions, i.e., POI visits. The analyst wishes to understand what visitors do in the park by extracting and categorizing existing behavior patterns.

The analyst starts with an overview of the action probability distribution in all topics (Fig. 12a). The exploration process can follow different workflows. One option is to use filtering for considering separately the widespread actions

(Fig. 12c) and more specific topic-related actions (Fig. 12b). The analyst can observe that a majority of the visitors attended “rides for everyone”, “shopping”, “food” and “restroom”. This observation corresponds to the common sense and thus increases the analyst’s trust to the data and results. Another option is to use the topic projection (Fig. 12d,e) and identify two distinct groups. The first group represents visitors addicted to “thrill rides” (Fig. 12f) and the second those who prefer relaxing activities, e.g. “shopping”, “kiddie rides”, “rides for everyone”, etc.

The analyst continues the exploration of these two topic groups (Fig. 13) and refines them by progressive application of group editing functions. Thus, thrill rider behaviors were further divided into four more specific topic groups. For example, G2 focuses on specific thrill rides “Atmosfear and Galactasaurus” (Fig. 13b). The analyst extracts a smaller group G1 which represents shopping behaviors in addition to the thrill rides (Fig. 13c). Group G3 includes one “thrill ride”, two “rides for everyone”, one “shopping” and one “food” (Fig. 13d), which may be interpreted as a touristic pattern with no strong preference for any attraction. Another small group of topics includes “beer” and “thrill rides” (Fig. 13e). In a similar way, the analyst explores in detail and refines the second of the two initial topic groups, which is focused on relaxing actions. The analyst identifies subgroups focused on “kiddie rides” and “shopping” (Fig. 13g), and “beer” combined with “shopping” and “rides for everyone” (Fig. 13f). During the progressive refinement process, group G12 was removed due to its large overlap with another group.

The analysis ends up in creating 11 topic groups whose medoids are treated as representations (models) of behavior categories. The matrix view in Fig. 14 shows that each topic is dominated by specific actions, and the topics have quite small overlaps, which makes even the Venn diagram applicable in this case. The results of the analysis reveal typical patterns of user behaviors. This can provide a basis for detection of unusual events in the IEEE VAST Challenge 2015 [77]; however, complete solution of the Challenge is beyond the scope of this work.

7 DISCUSSION AND CONCLUSION

We presented an approach to exploration and categorization of behaviors employing a combination of topic modeling with visual analytics techniques. The proposed workflow conforms to the view of visual analytics as a process of building a model of some subject through iterative evaluation and progressive development of the current model until it satisfies appropriateness criteria [78]. Thus, we explicitly consider the criteria that are relevant for evaluating a behavioral topic model and provide tools supporting the evaluation and further development.

Our approach does not rely on setting a fixed target number of topics, which is often unknown. As a special case, our approach also supports the traditional single-round LDA enabling interactive visual exploration of results. The benefit of executing multiple runs of LDA is obtaining a new information layer with diverse topics of varied specificity for further analysis. Strong behavior patterns (such as the thrill riding behaviors in Case 2) become prominent due

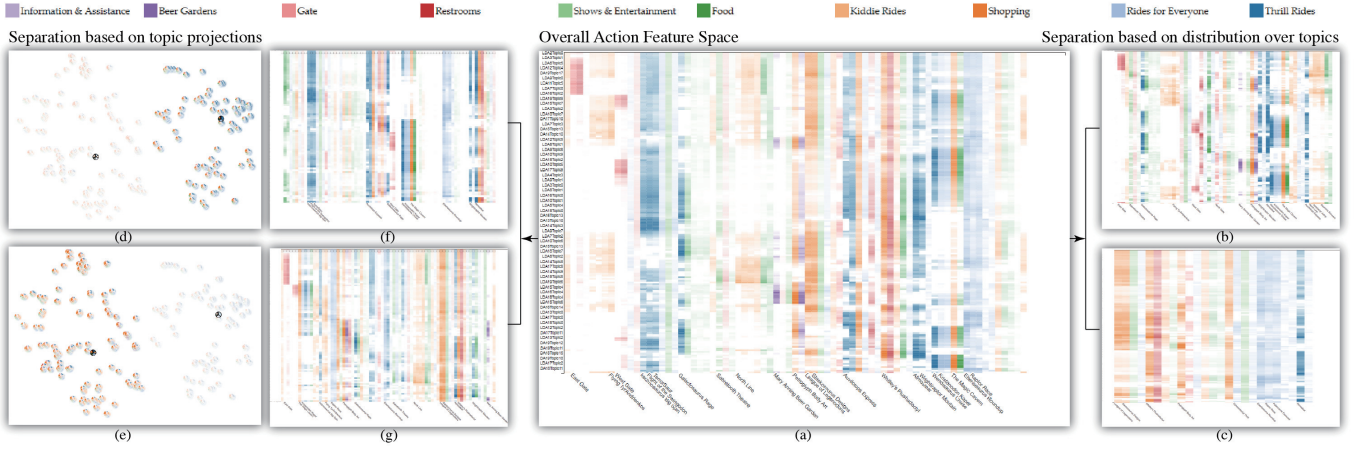


Fig. 12. Initial pattern exploration in visiting behaviors. (a) Overall actions distribution over the topics. (b)-(d) Different subsets of behaviors selected based on action probabilities distributions. (f)-(g) Subsets of behaviors based on topic grouping.

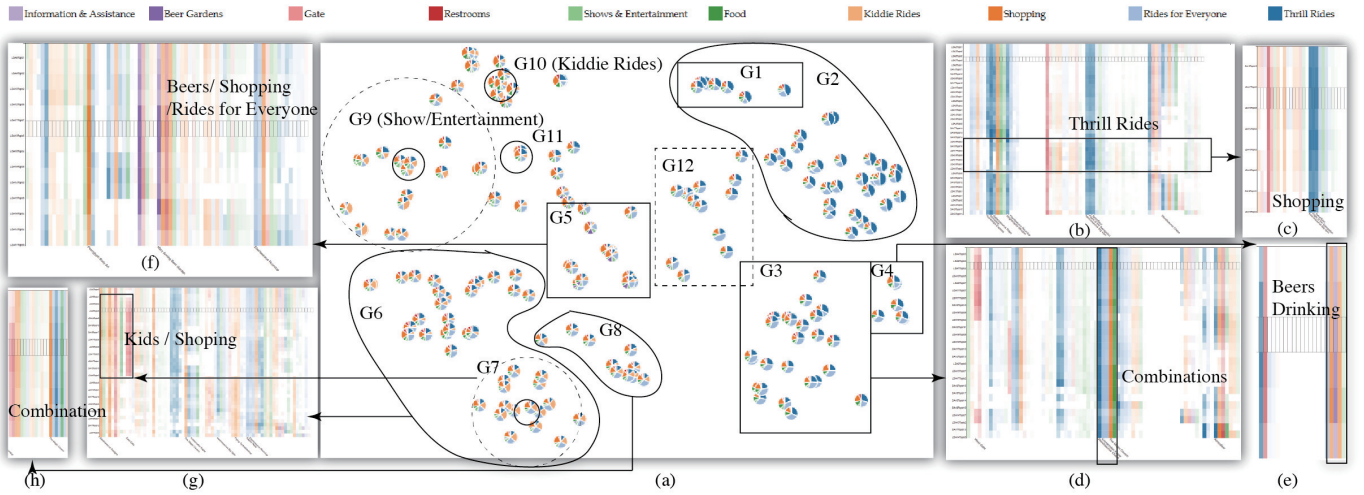


Fig. 13. Topic grouping for categorizing user visiting behaviors. (a) Creation of topic groups in the topic projection. (b)-(e) Groups of topics containing mostly thrill rides. (f)-(h) Groups of topics with more relaxing activities, such as kiddie rides, shopping, and rides for everyone.

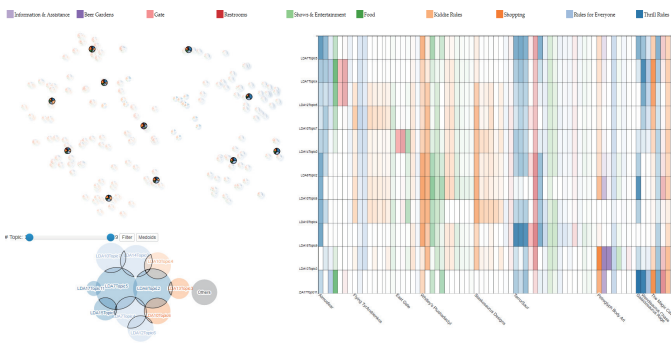


Fig. 14. The final analysis result consists of 11 topics.

to the existence of multiple similar topics resulting from different runs. Subtler behavior patterns (such as the beer drinking combined with thrill riding in Case 2), which might only be identified by some runs of LDA, can be found through the interactive exploration process. By including more general or more specific topics in the final model, the analyst can accommodate its level of discriminability according to application needs.

The proposed approach helps analysts not only to iden-

tify behavior types but also to understand the feature space and make informed decisions that influence the overall behavior categorization process. The idea of progressive topic exploration and selection advances the traditional use of the topic modeling methods.

One of the limitations of the approach is that it does not fully remove the need to estimate the number of topics, as it requires to set the upper bound. This decision can be facilitated by inspecting the coverage of the generated topics and using the cross-verification method [79]. The scalability with respect to the number of topics may also be an issue. To handle it, we envision two potentially useful approaches: sampling for reducing the number of topics and topic similarity checks (e.g. [15]) for filtering out similar topics.

Since our approach disregards the sequence in which actions occur, it is not suitable for tasks where action order is of primary importance. As explained in Section 2, the purpose of our approach is different from and complementary to sequence analysis. For gaining more comprehensive understanding of behaviors, it can be worthwhile to apply both behavior topic modelling and sequence analysis supported by interactive visual tools (e.g., [5], [6]).

We demonstrate the validity of the proposed approach through applied case studies in two substantially different application domains. We can envision further application areas, such as analysis of shopping behaviors with the aim to provide recommendations to customers or predict demands.

We plan two directions of further research and development. First, we want to find and incorporate suitable methods for behavior prediction, further combining automatic calculation and interactions. Second, we aim to consider more sophisticated definitions of behaviors taking into account temporal, sequential, and quantitative information and thus develop further approaches to behavior analysis.

In respect to the computer security application domain, we are going to address suggestions and follow-up requirements of our experts. Particularly, they noted that the results of our analysis are helpful but not sufficient for fraud detection. The value of our approach is that it is capable to identify typical and atypical behaviors and provide deep understanding of the distribution of behavior types and their representative actions. Such understanding can help and trigger domain experts to further develop and improve their existing model. Still, it appears possible to extend our approach for being more directly suited to the task of fraud detection. This requires using additional information concerning the sessions (time, IP address, operation platform, etc.) and the users (history of previous activities, typical session settings, etc.). Hence, behavior topic modeling needs to be combined with techniques for making inferences based on these kinds of information.

During the evaluation of the resulting behaviors, one emerging requirement was to closely inspect the behaviors extracted and relate individual sessions to these models. We currently made use of flow maps but further work is needed to better integrate such detailed representations within our approach to support the evaluation of the topics.

In the reported study, domain experts were involved in the evaluation of the approach but not of the visual analytics system implementing it. We intend to conduct a follow-on user study in which the experts will operate the system by themselves and evaluate its usability.

ACKNOWLEDGMENTS

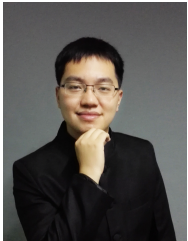
This research was supported by Fraunhofer Cluster of Excellence on “Cognitive Internet Technologies”, by EU in projects DiSIEM and SoBigData and by DFG (German Research Foundation) in priority research program SPP 1894 “Volunteered Geographic Information: Interpretation, Visualization and Social Computing”. The authors are grateful to M. Kamp, G. Fuchs and J. Sicking (Fraunhofer IAIS) and to project partners for discussions and collaboration.

REFERENCES

- [1] T. Fawcett and F. Provost, “Activity monitoring: Noticing interesting changes in behavior,” in *ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1999, pp. 53–62.
- [2] S. Chen, L. Lin, and X. Yuan, “Social Media Visual Analytics,” *Computer Graphics Forum*, vol. 36, no. 3, pp. 563–587, 2017.
- [3] D. M. Blei and J. D. Lafferty, “Topic models,” in *Text Mining*. Chapman and Hall/CRC, 2009, pp. 101–124.
- [4] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, “Mining user similarity based on location history,” in *Proc. ACM SIGSPATIAL GIS*, 2008, pp. 34:1–34:10.
- [5] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman, “Temporal event sequence simplification,” *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2227–2236, 2013.
- [6] P. H. Nguyen, C. Turkey, G. Andrienko, N. Andrienko, O. Thonnard, and J. Zouaoui, “Understanding user behaviour through action sequences: from the usual to the unusual,” *IEEE transactions on visualization and computer graphics*, 2018.
- [7] S. Guo, Z. Jin, D. Gotz, F. Du, H. Zha, and N. Cao, “Visual progression analysis of event sequence data,” *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2018.
- [8] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information processing & management*, vol. 24, no. 5, pp. 513–523, 1988.
- [9] M. Steinbach, L. Ertz, and V. Kumar, “The challenges of clustering high-dimensional data,” in *In New Vistas in Statistical Physics: Applications in Econophysics, Bioinformatics, and Pattern Recognition*. Springer-Verlag, 2003.
- [10] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.
- [11] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, “Indexing by latent semantic analysis,” *Journal of the American society for information science*, vol. 41, no. 6, pp. 391–407, 1990.
- [12] T. Hofmann, “Probabilistic latent semantic analysis,” in *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, ser. UAI’99. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999, pp. 289–296.
- [13] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [14] T. Rattenbury and J. Canny, “CAAD: An automatic task support system,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’07. New York, NY, USA: ACM, 2007, pp. 687–696.
- [15] J. Chuang, M. E. Roberts, B. M. Stewart, R. Weiss, D. Tingley, J. Grimmer, and J. Heer, “Topiccheck: Interactive alignment for assessing topic model stability,” in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015, pp. 175–184.
- [16] R. Arun, V. Suresh, C. E. Veni Madhavan, and M. N. Narasimha Murthy, “On finding the natural number of topics with latent dirichlet allocation: Some observations,” in *Advances in Knowledge Discovery and Data Mining*, M. J. Zaki, J. X. Yu, B. Ravindran, and V. Pudi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 391–402.
- [17] J. Yin and J. Wang, “A dirichlet multinomial mixture model-based approach for short text clustering,” in *Proceedings of the 20th ACM SIGKDD*. ACM, 2014, pp. 233–242.
- [18] D. Blei, T. Griffiths, M. Jordan, and J. Tenenbaum, “Hierarchical topic models and the nested chinese restaurant process,” *Advances in neural information processing systems*, vol. 16, pp. 106–114, 2004.
- [19] L. Liu, L. Tang, L. He, W. Zhou, and S. Yao, “An overview of hierarchical topic modeling,” *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, vol. 01, pp. 391–394, 2016.
- [20] D. Mimno, W. Li, and A. McCallum, “Mixtures of hierarchical topics with pachinko allocation,” in *Proceedings of the 24th International Conference on Machine Learning*, ser. ICML ’07. New York, NY, USA: ACM, 2007, pp. 633–640. [Online]. Available: <http://doi.acm.org/10.1145/1273496.1273576>
- [21] J. Chen, J. Zhu, J. Lu, and S. Liu, “Scalable training of hierarchical topic models,” *Proc. VLDB Endow.*, vol. 11, no. 7, pp. 826–839, Mar. 2018.
- [22] G. Salton and M. J. McGill, “Introduction to modern information retrieval,” 1986.
- [23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [24] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda, “Topic tracking model for analyzing consumer purchase behavior,” in *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, ser. IJCAI’09. Morgan Kaufmann Publishers Inc., 2009, pp. 1427–1432.
- [25] J. Zhu, A. Ahmed, and E. P. Xing, “Medlda: Maximum margin supervised topic models for regression and classification,” in

- Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. ACM, 2009, pp. 1257–1264.
- [26] K. Farrahi and D. Gatica-Perez, "Discovering routines from large-scale human locations using probabilistic topic models," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 1, pp. 3:1–3:27, Jan. 2011.
- [27] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, ser. EMNLP '09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 248–256.
- [28] D. M. Blei and J. D. McAuliffe, "Supervised topic models," *arXiv preprint arXiv:1003.0783*, 2010.
- [29] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06. New York, NY, USA: ACM, 2006, pp. 113–120.
- [30] T. Y. Lee, A. Smith, K. Seppi, N. Elmqvist, J. Boyd-Graber, and L. Findlater, "The human touch: How non-expert users perceive, interpret, and fix topic models," *International Journal of Human-Computer Studies*, vol. 105, no. Supplement C, pp. 28 – 42, 2017.
- [31] M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi, "Eddi: interactive topic-based browsing of social status streams," in *Proceedings of the 23rd annual ACM symposium on User interface software and technology*. ACM, pp. 303–312.
- [32] J. Chuang, C. D. Manning, and J. Heer, "Termite: Visualization techniques for assessing textual topic models," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 2012, pp. 74–77.
- [33] E. Alexander, J. Kohlmann, R. Valenza, M. Witmore, and M. Gleicher, "Serendip: Topic model-driven visual exploration of text corpora," in *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2014, pp. 173–182.
- [34] S. Rönqvist, A. Wang, and P. Sarlin, "Interactive visual exploration of topic models using graphs," in *The Eurographics Conference on Visualization (EuroVis)*, R. S. Laramée and M. Chen, Eds. Eurographics, 2014.
- [35] Y.-R. Lin, J. Sun, N. Cao, and S. Liu, "Contextour: Contextual contour visual analysis on dynamic multi-relational clustering," in *Proceedings of the 2010 SIAM International Conference on Data Mining*. SIAM, 2010, pp. 418–429.
- [36] N. Cao, J. Sun, Y.-R. Lin, D. Gotz, S. Liu, and H. Qu, "FacetAtlas: Multifaceted visualization for rich text corpora," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 1172–1181, 2010.
- [37] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang, "Tiara: a visual exploratory text analytic system," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 153–162.
- [38] M. Dork, D. Gruen, C. Williamson, and S. Carpendale, "A visual backchannel for large-scale events," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 1129–1138, 2010.
- [39] S. Havre, E. Hetzler, P. Whitney, and L. Nowell, "Themeriver: Visualizing thematic changes in large document collections," *IEEE transactions on visualization and computer graphics*, vol. 8, no. 1, pp. 9–20, 2002.
- [40] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. J. Gao, H. Qu, and X. Tong, "Textflow: Towards better understanding of evolving topics in text," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2412–2421, 2011.
- [41] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky, "Hierarchical-topics: Visually exploring large text collections using topic hierarchies," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2002–2011, 2013.
- [42] W. Cui, S. Liu, Z. Wu, and H. Wei, "How hierarchical topics evolve in large text corpora," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2281–2290, 2014.
- [43] X. Wang, S. Liu, J. Liu, J. Chen, J. Zhu, and B. Guo, "Topic-panorama: A full picture of relevant topics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 12, pp. 2508–2521, Dec 2016.
- [44] E. Alexander and M. Gleicher, "Task-driven comparison of topic models," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 320–329, 2016.
- [45] J. Chuang, S. Gupta, C. D. Manning, and J. Heer, "Topic model diagnostics: Assessing domain relevance via topical alignment," in *International Conference on Machine Learning (ICML)*, 2013.
- [46] H. Lee, J. Kihm, J. Choo, J. Stasko, and H. Park, "ivisclustering: An interactive visual document clustering via topic modeling," *Computer Graphics Forum*, vol. 31, no. 3pt3, pp. 1155–1164, 2012.
- [47] E. Hoque and G. Carenini, "Convisit: Interactive topic modeling for exploring asynchronous online conversations," in *Proceedings of the 20th International Conference on Intelligent User Interfaces*, ser. IUI '15. New York, NY, USA: ACM, 2015, pp. 169–180.
- [48] J. G. S. Paiva, W. R. Schwartz, H. Pedrini, and R. Minghim, "An approach to supporting incremental visual data classification," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 1, pp. 4–17, Jan 2015.
- [49] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of artificial intelligence research*, 1996.
- [50] J. Choo, C. Lee, C. K. Reddy, and H. Park, "UTOPIAN: User-driven topic modeling based on interactive nonnegative matrix factorization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 1992–2001, 2013.
- [51] M. Kim, K. Kang, D. Park, J. Choo, and N. Elmqvist, "TopicLens: Efficient multi-level visual topic exploration of large-scale document collections," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 151–160, Jan 2017.
- [52] M. El-Assady, R. Sevastjanova, F. Sperrle, D. Keim, and C. Collins, "Progressive learning of topic modeling parameters: A visual analytics framework," *IEEE Transactions on Visualization and Computer Graphics*, vol. PP, no. 99, pp. 1–1, 2017.
- [53] D. Andrzejewski, X. Zhu, and M. Craven, "Incorporating domain knowledge into topic modeling via dirichlet forest priors," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: ACM, 2009, pp. 25–32. [Online]. Available: <http://doi.acm.org/10.1145/1553374.1553378>
- [54] Y. Hu, J. Boyd-Graber, B. Satinoff, and A. Smith, "Interactive topic modeling," *Machine learning*, vol. 95, no. 3, pp. 423–469, 2014.
- [55] N. Andrienko and G. Andrienko, "State transition graphs for semantic analysis of movement behaviours," *Information Visualization*, vol. 17, no. 1, pp. 41–65, 2018.
- [56] P. Nguyen, C. Turkay, G. Andrienko, N. Andrienko, and O. Thonnard, "A visual analytics approach for user behaviour understanding through action sequence analysis," in *EuroVis Workshop on Visual Analytics*, 2017.
- [57] R. Krüger, S. Lohmann, D. Thom, H. Bosch, and T. Ertl, "Using social media content in the visual analysis of movement data," in *Proc. Workshop on Interactive Visual Text Analytics*, 2012.
- [58] T. Kanda, M. Shiomi, L. Perrin, T. Nomura, H. Ishiguro, and N. Hagita, "Analysis of people trajectories with ubiquitous sensors in a science museum," in *Proc. of IEEE ICRA*, 2007, pp. 4846–4853.
- [59] R. Xiong and J. Donath, "Peoplelegend: creating data portraits for users," in *Proc. of the ACM UIST*, 1999, pp. 37–44.
- [60] D. Orellana, M. Wachowicz, N. Andrienko, and G. Andrienko, "Uncovering interaction patterns in mobile outdoor gaming," in *Proc. GEOProcessing*, 2009, pp. 177–182.
- [61] S. Koldijk, J. Bernard, T. Ruppert, J. Kohlhammer, M. Neerincx, and W. Kraaij, "Visual Analytics of Work Behavior Data - Insights on Individual Differences," in *Eurographics Conference on Visualization (EuroVis) - Short Papers*, E. Bertini, J. Kennedy, and E. Puppo, Eds. The Eurographics Association, 2015.
- [62] T. Blascheck, F. Beck, S. Baltes, T. Ertl, and D. Weiskopf, "Visual analysis and coding of data-rich user behavior," in *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, Oct 2016, pp. 141–150.
- [63] J. Peng, K.-K. R. Choo, and H. Ashman, "User profiling in intrusion detection: A review," *Journal of Network and Computer Applications*, vol. 72, pp. 14–27, 2016.
- [64] G. Pannell and H. Ashman, "User modelling for exclusion and anomaly detection: a behavioural intrusion detection system," in *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 2010, pp. 207–218.
- [65] M. Whiting, K. Cook, G. Grinstein, J. Fallon, K. Liggett, D. Staheli, and J. Crouser, "Vast challenge 2015: Mayhem at dinofun world," in *IEEE Conference on Visual Analytics Science and Technology (VAST) 2015*. IEEE, 2015, pp. 113–118.
- [66] G. Heinrich, "Parameter estimation for text analysis," Fraunhofer IGD, Tech. Rep., 2009.
- [67] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

- [68] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [69] C. Sievert and K. Shirley, "Ldavis: A method for visualizing and interpreting topics," in *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, 2014, pp. 63–70.
- [70] "https://shuauiw.github.io/2016/12/22/topic-modeling-and-tsne-visualization.html, last accessed 2018/09/21."
- [71] P. E. Black, "Manhattan distance," *Dictionary of Algorithms and Data Structures*, vol. 18, p. 2012, 2006.
- [72] S. Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, 2nd ed. USA: Analytics Press, 2012.
- [73] J. W. Tukey, "Exploratory data analysis," 1977.
- [74] J. Caldas, N. Gehlenborg, A. Faisal, A. Brazma, and S. Kaski, "Probabilistic retrieval and visualization of biologically relevant microarray experiments," *BMC bioinformatics*, vol. 10, no. 13, 2009.
- [75] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel, *Visual Analytics of Movement*. Springer, 2013.
- [76] N. Andrienko, G. Andrienko, G. Fuchs, and P. Jankowski, "Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces," *Information Visualization*, vol. 15, no. 2, pp. 117–153, 2016.
- [77] "IEEE VAST challenge 2015 benchmark. <https://www.cs.umd.edu/hcil/varepository/vast>, last accessed 09/21/2018."
- [78] N. Andrienko, T. Lammarsch, G. Andrienko, G. Fuchs, D. Keim, S. Miksch, and A. Rind, "Viewing visual analytics as model building," *Computer Graphics Forum*, vol. 37, no. 6, pp. 275–299, 2018.
- [79] K. Hornik and B. Grün, "topicmodels: An r package for fitting topic models," *Journal of Statistical Software*, vol. 40, no. 13, pp. 1–30, 2011.



Siming Chen is a research scientist at Fraunhofer Institute IAIS and a PostDoc researcher of University of Bonn in Germany. He got his PhD in Peking University. His research interests include visual analytics of social media, cyber security and spatial temporal data. He published several papers in IEEE VIS, IEEE TVCG, EuroVis, etc. More information can be found in <http://simingchen.me>.



Natalia Andrienko is a lead scientist responsible for visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems and part-time professor at City University London. Results of her research have been published in two monographs "*Exploratory Analysis of Spatial and Temporal Data: a Systematic Approach*" (Springer 2006) and "*Visual Analytics of Movement*" (Springer 2013). Natalia Andrienko is an associate editor of *IEEE Transactions on Visualization and Computer Graphics*.



Gennady Andrienko is a lead scientist responsible for visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems and part-time professor at City University London. Gennady Andrienko was a paper chair of *IEEE VAST* conference (2015–2016) and associate editor of *IEEE Transactions on Visualization and Computer Graphics* (2012–2016), *Information Visualization* and *International Journal of Cartography*.



Linara Adilova is currently employed in Fraunhofer IAIS, Knowledge Discovery department as a researcher. She graduated from University of Bonn with a Master of Science degree for Computer Science in 2017. Her main interests are Deep Learning and Natural Language Processing.



Jeremie Barlet is a software development manager for the Amadeus IT Group, the leading provider of IT solutions to the global tourism and travel industry. His is leading a development team in charge of the Fraud detection platform which leverages on machine learning and user behaviour analytics to detect fraudulent activities in Amadeus applications. After 5 years heading the Logon and Security Server development team, he could use his functional expertise to contribute to the project.



Jörg Kindermann is a mathematician and computer linguist. He graduated from the University of Bielefeld in 1986. From 1987 to 2001 he worked at GMD (Gesellschaft für Mathematik und Datenverarbeitung). Since 2001 he is a senior research scientist at the Fraunhofer Institute IAIS. His research interests include text mining, neural networks, machine learning algorithms, and high performance computing.



Phong H. Nguyen is a Research Associate at the giCentre, City, University of London. His research mainly focuses on the design and application of interactive visualizations to make sense of complex datasets, with a special interest in analytic provenance, logs and general temporal categorical data. He has published papers in high-impact journals including IEEE TVCG, InfoVis, VAST, CG&A and IVS. Phong holds a PhD in Visual Analytics from Middlesex University, London, UK.



Olivier Thonnard is a Senior Security Expert for the Amadeus IT Group, the leading provider of IT solutions to the global tourism and travel industry. His R&D activities focus on machine learning and information visualization for real-world security applications, especially for fraud detection based on User Behaviour Analytics. Dr Thonnard was a Senior Principal Research Engineer at Symantec Research Labs, focusing on technology innovation and thought leadership in computer and network security.



Catagay Turkey is a Senior Lecturer in Applied Data Science at the giCentre at the Computer Science Department of City, University of London conducting research in visual analytics. He serves as a committee member for several conferences including InfoVis and EuroVis, and part of the organising committee for IEEE VIS on 2017 and 2018. He is currently a guest editor for *IEEE Computer Graphics and Applications*, and an editorial board member for the *Machine Learning and Knowledge Extraction* journal.